ABSTRACT OF THE DISCLOSURE

SYSTEM AND METHOD FOR DETECTING EXTENSIBLE PATTERS

Given an input sequence of data, a rigid pattern is a repeating sequence, possibly interspersed with don't-care characters. The data can be a sequence of characters or sets of characters or even real values. In practice, the patterns or motifs of interest are the ones that also allow a variable number of gaps (or don't-care characters): these are patterns with spacers termed extensible patterns. In a bioinformatics context, similar patterns have also been called flexible patterns or motifs. A system according to the invention discovers all the maximal extensible motifs in the input. The flexibility is succinctly defined by a single integer parameter $D \geq 1$ which is interpreted as the allowable space to be between 1 and D characters between two successive solid characters in a reported motif.